Blind Identification Algorithms for Sparse SIMO Systems

Abdeldjalil Aissa El Bey

Blind Identification Algorithms for Sparse SIMO Systems

Abdeldjalil Aïssa-El-Bey

Institut Télécom; Télécom Bretagne; UMR CNRS 6285 Lab-STICC, Technopole Brest Iroise CS 83818 29238 Brest, France

Université Européenne de Bretagne, France

E-mail: abdeldjalil.aissaelbey@telecom-bretagne.eu

Abstract

In this work, we are interested in blind identification of sparse single-input multpe-output (SIMO) systems. First, we present a modified cross-relations (CR) technique combined with a ℓ_p norm, which is considered as a good sparsity measure. Then, A maximum a posteriori approach is considered using generalized Laplacian distribution for the channel coefficients. This leads to a cost function given by the deterministic maximum likelihood (ML) criterion penalized by 'a sparsity measure' term expressed by the ℓ_p norm of the channel coefficient vector. A simple but efficient optimization algorithm using gradient technique with optimal step-size is proposed. The simulations show that the proposed method outperforms the ML technique in terms of estimation error and is robust against channel order overestimation errors.

Index Terms

Blind system identification, SIMO, Sparse.

I. Introduction

Blind system identification (BSI) is a fundamental signal processing technology aimed at retrieving a system's unknown information from its outputs only. This problem has received a lot of attention in

the signal processing literature and a plethora of methods and techniques have been proposed to solve the BSI over the last two decades [1]–[3]. Techniques for BSI can generally be classified into two main classes (i) higher order statistical (HOS) and (ii) second order statistical (SOS) methods. Although HOS methods [1] were proposed for BSI due to the rich information, large number of observation samples are required. As a result, SOS methods such as [4] have become more popular. Comparison between SOS and HOS methods have been presented in [3]. Unfortunately, these methods have demonstrated their limitation when channel impulse response is very long and sparse (e.g. HF communication, echo cancelation, etc).

Estimation of sparse long channels (i.e. channels with small number of nonzero coefficients but a large span of delays) is considered in this paper. Such sparse channels are encountered in many communication applications: High-Definition television (HDTV) channels are hundreds of data symbols long but there are only a few nonzero taps [5]. Hilly terrain delay profile has a small number of multipath in the broadband wireless communication [6] and underwater acoustic channels are also known to be sparse [7]–[9]. As such, underwater acoustic channels typically exhibit significant delay spreads, but with very few dominant multipath components [10]–[12].

In this paper we propose, to exploit the sparse nature of the channel impulse response via a maximum a posteriori (MAP) approach using a Generalized Gaussian Distribution to model the sparse channel type. As will be shown in the sequel, the MAP criterion combines the maximum-likelihood (ML) cost function with the ℓ_p norm constraint ($0 \leq p \leq 1$) of the channel impulse response, which is considered by many authors as a good sparsity measure, e.g. [13], [14].

In the following section, we discuss the data model that formulates our problem. Next, we review the cross-relations (CR), deterministic ML and channel subspace (CS) methods for blind SIMO channel identification before using it to introduce the Iterative Sparse Blind System Identification (ISBSI) and MAP solutions. In Section VI, some simulations are undertaken to validate our algorithm and illustrate

its robustness against channel overestimation errors and to compare its performance to other existing BSI techniques.

II. Problem Formulation

The problem addressed in this paper is to determine the sparse impulse response of a SIMO system in a blind way, i.e. only the observed system outputs are available and used without assuming knowledge of the specific input signal.

Consider a mathematical model where the input and the output are discrete, the system is driven by a single-input sequence $s(n)$ and yields M output sequences $x_1(n), \ldots, x_M(n)$, and the system has finite impulse responses (FIR's) $h_i(n) \neq 0$, for $n = 0, \ldots, L$ and $i = 1, \ldots, M$. Such a system model can be described as follows :

$$\begin{cases} x_1(n) &= s(n) * h_1(n) &+ w_1(n) \\ x_2(n) &= s(n) * h_2(n) &+ w_2(n) \\ &\vdots & \\ x_M(n) &= s(n) * h_M(n) &+ w_M(n) \end{cases} \tag{1}$$

where $*$ denotes linear convolution and $\boldsymbol{w}(n) = [w_1(n), \ldots, w_M(n)]^T$ is an additive spatial white noise, i.e. $\mathbb{E}[\boldsymbol{w}(n)\boldsymbol{w}(n)^H] = \sigma^2 \boldsymbol{I}_M$ where $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and the conjugate transpose, respectively and $\boldsymbol{I}_M$ is a $M \times M$ identity matrix. In vector form, equation (1) can be expressed as :

$$\boldsymbol{x}(n) = \sum_{k=0}^{L} \boldsymbol{h}(k) s(n-k) + \boldsymbol{w}(n) \ ,$$

where $\boldsymbol{h}(z) = \sum_{k=0}^{L} \boldsymbol{h}(k) z^{-k}$ is an unknown causal FIR $M \times 1$ transfer function satisfying $\boldsymbol{h}(z) \neq 0, \ \forall z$. Given a finite set of observation vectors $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(T)$ and based on the channel entries co-primness (i.e. $\boldsymbol{h}(z) \neq 0 \ \forall z$), the objective here is to estimate the channel coefficients vector $\boldsymbol{h} = [\boldsymbol{h}(0)^T, \cdots, \boldsymbol{h}(L)^T]^T$ up to a scalar constant (this is an inherent indeterminacy of the blind system identification problem as shown in [4]).

III. IDENTIFICATION SIMO SYSTEMS

In this section, we introduce a new methods for blind SIMO systems identification based on sparse assumption of channel impulse response. Therefore, we start this section by a brief overview of well known maximum-likelihood (ML) method [2], in order to facilitate the introduction of the sparsity assumption based approach via maximum a posteriori (MAP) probability method and its variants.

A. Maximum Likelihood Method

maximum-likelihood (ML) is a classic approach applicable to any parameter estimation problem where the probability density function (PDF) of the available data is known. Assuming that the system output vector is corrupted by additive white Gaussian noise vector, the system output vector can be rewrite as

$$\boldsymbol{x} = \boldsymbol{\mathcal{H}}_M \boldsymbol{s} + \boldsymbol{w} \tag{2}$$

and the PDF of $\boldsymbol{x}$ is given by

$$f(\boldsymbol{x}|\boldsymbol{h}) = \frac{1}{(2\pi)^{\frac{T}{2}}\sigma^T} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{x} - \boldsymbol{\mathcal{H}}_M \boldsymbol{s}\|_2^2\right)$$

where σ^2 is the variance of each element of $\boldsymbol{w}$. The ML estimates of $\boldsymbol{\mathcal{H}}_M$ and $\boldsymbol{s}$ are given by those arguments that maximize the PDF $f(\boldsymbol{x})$

$$(\boldsymbol{\mathcal{H}}_M, \boldsymbol{s}) = \arg\max_{\boldsymbol{\mathcal{H}}_M, \boldsymbol{s}} f(\boldsymbol{x}|\boldsymbol{h}) \tag{3}$$

$$= \arg\min_{\boldsymbol{\mathcal{H}}_M, \boldsymbol{s}} \left\{\|\boldsymbol{x} - \boldsymbol{\mathcal{H}}_M \boldsymbol{s}\|_2^2\right\} \tag{4}$$

where proper constraints on $\boldsymbol{\mathcal{H}}_M$ and $\boldsymbol{s}$ are imposed. Note that such ML criterion is equivalent to the least-square (LS) criterion, for which the knowledge of the PDF of $\boldsymbol{x}$ is not necessary. For any given $\boldsymbol{\mathcal{H}}_M$, the ML estimate of that minimizes the quadratic function $\|\boldsymbol{x} - \boldsymbol{\mathcal{H}}_M \boldsymbol{s}\|_2^2$ is known to be

$$\widehat{\boldsymbol{s}} = \left(\boldsymbol{\mathcal{H}}_M^H \boldsymbol{\mathcal{H}}_M\right)^{-1} \boldsymbol{\mathcal{H}}_M^H \boldsymbol{x}$$

(Under the necessary identifiability condition, the matrix $\boldsymbol{\mathcal{H}}_M$ is known to have full column rank [15].) Using this estimate in equation (4) yields

$$\widehat{\boldsymbol{\mathcal{H}}}_M = \arg\min_{\boldsymbol{\mathcal{H}}_M} \left\{ \|(\boldsymbol{I}_M - \boldsymbol{\mathcal{P}}_H)\boldsymbol{x}\|_2^2 \right\} \tag{5}$$

where $\boldsymbol{\mathcal{P}}_H$ is the orthogonal projection matrix onto the range of $\boldsymbol{\mathcal{H}}_M$, i.e.

$$\boldsymbol{\mathcal{P}}_H = \boldsymbol{\mathcal{H}}_M \left(\boldsymbol{\mathcal{H}}_M^H \boldsymbol{\mathcal{H}}_M \right)^{-1} \boldsymbol{\mathcal{H}}_M^H$$

Although the minimization in (5) is computationally much more efficient than that in (4), it is still highly nonlinear. Therefore, the computation of (5) has to be iterative in nature. Many iterative optimization approaches such as [16], [17] can be applied to compute (5). In this paper, we have chosen to use the two-step ML (TSML) approach presented in [2]. In this case the channel estimation is given by (for more details see [2]):

$$\widehat{\boldsymbol{h}} = \arg\min_{\boldsymbol{h}} \left\{ \boldsymbol{h}^H \boldsymbol{\mathcal{X}}_M^H \left(\boldsymbol{\mathcal{G}}_M^H \boldsymbol{\mathcal{G}}_M \right)^{\#} \boldsymbol{\mathcal{X}}_M \boldsymbol{h} \right\} \tag{6}$$

where the superscript $(\cdot)^{\#}$ denotes a Moore-Penrose pseudoinverse operator, $\boldsymbol{\mathcal{X}}_M$ is defined by:

$$\boldsymbol{\mathcal{X}}_2 = [\boldsymbol{X}_2, \ -\boldsymbol{X}_1] \tag{7}$$

and

$$\boldsymbol{\mathcal{X}}_l = \left[\begin{array}{ccc|c} & \boldsymbol{\mathcal{X}}_{l-1} & & \boldsymbol{0} \\ \hline \boldsymbol{X}_l & & \boldsymbol{0} & -\boldsymbol{X}_1 \\ & \ddots & & \vdots \\ \boldsymbol{0} & & \boldsymbol{X}_l & -\boldsymbol{X}_{l-1} \end{array}\right] \tag{8}$$

with $l = 3, \ldots, M$ and :

$$\boldsymbol{X}_l = \begin{bmatrix} x_l(L) & \ldots & x_l(0) \\ \vdots & & \vdots \\ x_l(T-1) & \ldots & x_l(T-L-1) \end{bmatrix}. \tag{9}$$

and $\boldsymbol{\mathcal{G}}_M$ is defined by:

$$\boldsymbol{\mathcal{G}}_2^H = \left[-\overline{\boldsymbol{H}}_1,\ \overline{\boldsymbol{H}}_2\right] \tag{10}$$

and

$$\boldsymbol{\mathcal{G}}_q^H = \left[\begin{array}{ccc|c} & \boldsymbol{\mathcal{G}}_{q-1}^H & & \mathbf{0} \\ \hline -\overline{\boldsymbol{H}}_q & & \mathbf{0} & \overline{\boldsymbol{H}}_1 \\ & \ddots & & \vdots \\ \mathbf{0} & & -\overline{\boldsymbol{H}}_q & \overline{\boldsymbol{H}}_{q-1} \end{array}\right] \tag{11}$$

where $q = 3, \ldots, M$ and $\overline{\boldsymbol{H}}_q$ is the top-left $(T-L) \times T$ submatrix of $\boldsymbol{H}_q$.

The expression (6) suggests the following TSML method.

- Step 1 : $\widehat{\boldsymbol{h}}_c = \underset{\|\boldsymbol{h}\|_2=1}{\arg\min} \left\{\boldsymbol{h}^H \boldsymbol{\mathcal{X}}_M^H \boldsymbol{\mathcal{X}}_M \boldsymbol{h}\right\}$
- Step 2 : $\widehat{\boldsymbol{h}}_e = \underset{\|\boldsymbol{h}\|_2=1}{\arg\min} \left\{\boldsymbol{h}^H \boldsymbol{\mathcal{X}}_M^H \left(\boldsymbol{\mathcal{G}}_c^H \boldsymbol{\mathcal{G}}_c\right)^{\#} \boldsymbol{\mathcal{X}}_M \boldsymbol{h}\right\}$, where $\boldsymbol{\mathcal{G}}_c$ is $\boldsymbol{\mathcal{G}}_M$ constructed from $\widehat{\boldsymbol{h}}_c$ according to equations (10) and (11)

The first step comes from (6) by setting the weighting matrix $\left(\boldsymbol{\mathcal{G}}_M^H \boldsymbol{\mathcal{G}}_M\right)^{\#}$ to an identity matrix. It can be shown that Step 1 of the algorithm yields the exact estimate of $\boldsymbol{h}$ in the absence of noise (or when the noise is white and the data length is infinite) and that Step 2 of the algorithm yields the optimum (ML) estimate of $\boldsymbol{h}$ at a relatively high signal-to-noise ratio (SNR).

B. Maximum a Posteriori Method

In this section, we introduce a Maximum a Posteriori (MAP) probability method which estimates the probability distribution of $\boldsymbol{h}$ as follows

$$\widehat{\boldsymbol{h}}_{MAP} = \arg\max_{\boldsymbol{h}} \left\{\frac{f(\boldsymbol{x}|\boldsymbol{h}) g(\boldsymbol{h})}{\int f(\boldsymbol{x}|\boldsymbol{h}') g(\boldsymbol{h}') d\boldsymbol{h}'}\right\} \tag{12}$$

$$= \arg\max_{\boldsymbol{h}} \left\{f(\boldsymbol{x}|\boldsymbol{h}) g(\boldsymbol{h})\right\} \tag{13}$$

To approximate the channel distribution in (13), we adopt the Generalized Gaussian Distribution (GGD) model, which can be mathematically represented under the assumption that all the component of $\boldsymbol{h}$ are *i.i.d*, as follows :

$$g(\boldsymbol{h}) = \left(\frac{p}{2\,\beta\,\Gamma\left(\frac{1}{p}\right)} \right)^{-M(L+1)} \exp\left(-\frac{\|\boldsymbol{h}\|_p^p}{\beta^p} \right) \tag{14}$$

where $\beta > 0$ is a scale parameter, $0 < p \leq 1$ and $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt,\ z > 0$, is the Gamma function. The new prior distribution gives more weight to values that are close to zero, thereby encouraging the model to set many latent variables to (or close to) zero. This makes it ideal for learning sparse representations.

The combination to equations (13) and (14) leads to the following objective function :

$$\mathcal{J}(\boldsymbol{h}) = \boldsymbol{h}^H \boldsymbol{\mathcal{X}}_M^H \left(\boldsymbol{\mathcal{G}}_c^H \boldsymbol{\mathcal{G}}_c \right)^{\#} \boldsymbol{\mathcal{X}}_M \boldsymbol{h} + \lambda \|\boldsymbol{h}\|_p^p \tag{15}$$

where $\lambda = \frac{2\sigma^2}{\beta^p}$ is a weighting parameter which controls the trade-off between approximation error and sparsity. The first term is the ML criterion and the second term is the penalty term, which minimizes the ℓ_p norm of the channel impulse response $\boldsymbol{h}$.

Therefor, the desired solution of $\boldsymbol{h}$ is determined by minimizing the cost function $\mathcal{J}(\boldsymbol{h})$ under the unit norm constraint $\|\boldsymbol{h}\|_2 = 1$:

$$\widehat{\boldsymbol{h}} = \arg \min_{\|\boldsymbol{h}\|_2 = 1} \left\{ \boldsymbol{h}^H \boldsymbol{\mathcal{X}}_M^H \left(\boldsymbol{\mathcal{G}}_c^H \boldsymbol{\mathcal{G}}_c \right)^{\#} \boldsymbol{\mathcal{X}}_M \boldsymbol{h} + \lambda \|\boldsymbol{h}\|_p^p \right\} \tag{16}$$

C. Sparse Cross-Relations Method

In order to simplify the optimization of the cost function given by (16), we propose to setting the weighting matrix $\left(\boldsymbol{\mathcal{G}}_M^H \boldsymbol{\mathcal{G}}_M \right)^{\#}$ to an identity matrix. This approximation is equivalent to the first step of the TSML method which yields the exact estimate of $\boldsymbol{h}$ in the absence of noise (or when the noise is white and the data length is infinite). Therefore, the desired solution of $\boldsymbol{h}$ is determined by minimizing

the cost function $\mathcal{J}(\boldsymbol{h})$ under the unit norm constraint $\|\boldsymbol{h}\|_2 = 1$:

$$\widehat{\boldsymbol{h}} = \arg \min_{\|\boldsymbol{h}\|_2=1} \left\{ \boldsymbol{h}^H \boldsymbol{\mathcal{X}}_M^H \boldsymbol{\mathcal{X}}_M \boldsymbol{h} + \lambda \|\boldsymbol{h}\|_p^p \right\} \ . \tag{17}$$

where the first term of equation (17) is equivalent to a cross-relations (CR) criterion [2], [4], [18], and the second term is the penalty term which minimizes the ℓ_p norm of the channel impulse response $\boldsymbol{h}$.

IV. Implementation

A. Gradient optimization

Direct minimization of the cost function given by (17) is computationally intensive and may be even intractable when the channel impulse response is long and when the number of channels is large. Here, a stochastic gradient technique is proposed to solve this minimization problem efficiently, and the solution is compute iteratively by :

$$\boldsymbol{h}_{k+1} = \boldsymbol{h}_k - \mu \nabla \mathcal{J}(\boldsymbol{h}_k) \ , \tag{18}$$

where μ is a small positive step size and ∇ is a gradient operator. The gradient of $\mathcal{J}(\boldsymbol{h})$ is given by :

$$\nabla \mathcal{J}(\boldsymbol{h}) = \frac{\partial \mathcal{J}(\boldsymbol{h})}{\partial \boldsymbol{h}} = 2 \ \boldsymbol{\mathcal{Q}}_M \boldsymbol{h} + \lambda \ \widetilde{\boldsymbol{h}} \ , \tag{19}$$

where

$$\widetilde{h}(i) = p \ \ \mathrm{sign}\left(h(i)\right) \ \ |h(i)|^{p-1} \qquad \text{for} \quad i = 1, \ldots, M(L+1) \ , \tag{20}$$

and $\boldsymbol{\mathcal{Q}}_M = \boldsymbol{\mathcal{X}}_M^H \boldsymbol{\mathcal{X}}_M$.

The unit norm constraint is to ensure that the iterative algorithm do not converge to a trivial solution with all zero elements. However, we observe that the gradient of the ℓ_p norm (20) may diverge if $|h(i)|$ is close to zero and $0 < p < 1$. Therefore, to avoid this problem, we introduce the parameter $\varepsilon > 0$ in order to provide stability and to ensure that a zero-valued component in $\boldsymbol{h}_k$ does not strictly prohibit a nonzero estimate at the next step

$$\widetilde{h}^{\varepsilon}(i) = p \ \ \mathrm{sign}\left(h(i)\right) \ \ (|h(i)| + \varepsilon)^{p-1} \qquad \text{for} \quad i = 1, \ldots, M(L+1) \ , \tag{21}$$

Therefore, the update equation is given by :

$$\boldsymbol{h}_{k+1} = \frac{\boldsymbol{h}_k - \mu \left(2\ \boldsymbol{\mathcal{Q}}_M \boldsymbol{h}_k + \lambda\ \widetilde{\boldsymbol{h}}_k^{\varepsilon}\right)}{\left\| \boldsymbol{h}_k - \mu \left(2\ \boldsymbol{\mathcal{Q}}_M \boldsymbol{h}_k + \lambda\ \widetilde{\boldsymbol{h}}_k^{\varepsilon}\right) \right\|_2} \ .$$

B. Optimal step size

In order to avoid divergence, a conservatively small μ is usually used, which inevitably sacrifies the convergence speed of the algorithm. In this section, we will derive an optimal step size for the gradient optimization method and hence propose a variable step size algorithm.

To find an optimal step size μ for each iteration we propose to use a line search method. More precisely, we choose a line search, in which μ is chosen to minimize $\mathcal{J}$

$$\mu = \arg\min_{\mu} \left\{ \mathcal{J}\left(\boldsymbol{h} - \mu \nabla \mathcal{J}(\boldsymbol{h})\right) \right\} \ . \tag{22}$$

The criterion in the $(k+1)^{th}$ iteration is written as follow :

$$\mathcal{J}(\boldsymbol{h}_{k+1}) = \|\boldsymbol{\mathcal{X}}_M \boldsymbol{h}_{k+1}\|_2^2 + \lambda \|\boldsymbol{h}_{k+1}\|_p^p \tag{23}$$

by replacing $\boldsymbol{h}_{k+1}$ by (18) we rewrite equation (23) as :

$$\mathcal{J}(\boldsymbol{h}_{k+1}) = \|\boldsymbol{\mathcal{X}}_M \left(\boldsymbol{h}_k - \mu \nabla \mathcal{J}(\boldsymbol{h}_k)\right)\|_2^2 + \lambda \left\|\boldsymbol{h}_k - \mu \nabla \mathcal{J}(\boldsymbol{h}_k)\right\|_p^p$$

we take a derivative of $\mathcal{J}(\boldsymbol{h}_{k+1})$ with respect to μ :

$$\frac{\partial \mathcal{J}(\boldsymbol{h}_{k+1})}{\partial \mu} = \mathcal{F}(\mu) = \left[\mu (2 \nabla \mathcal{J}(\boldsymbol{h}_k)^H \boldsymbol{\mathcal{Q}}_M - \lambda\ \widetilde{\boldsymbol{r}}_k^H) - 2 \boldsymbol{h}_k^H \boldsymbol{\mathcal{Q}}_M\right] \nabla \mathcal{J}(\boldsymbol{h}_k) \ ,$$

where

$$\widetilde{r}(i) = p\ sign\left(h(i) - \mu \nabla \mathcal{J}(\boldsymbol{h})(i)\right) |h(i) - \mu \nabla \mathcal{J}(\boldsymbol{h})(i)|^{p-1} \qquad \text{for} \quad i = 1, \ldots, M(L+1) \ .$$

Therefore, the optimal step size in each iteration is obtained in the form :

$$\mu_k = \mu_{k-1} - \mathcal{F}(\mu_{k-1}) \frac{\mu_{k-1} - \mu_{k-2}}{\mathcal{F}(\mu_{k-1}) - \mathcal{F}(\mu_{k-2})} \ ,$$

where we use an approximate Newton approach for solving (22).

C. Newton Optimization

In the previous section, an implementation based on stochastic gradient technique has been presented. While the algorithm has been shown to converge in the mean to the desired channel impulse responses, one of the difficulties in the design and implementation of the stochastic gradient is the selection of the step size μ. Aiming to achieve a good balance of design objectives, we present here a Newton method with variable step size during iteration.

Newton method is an efficient tool of optimization. It often converges fast and provides quadratic rate of convergence. However, its iteration may be costly, because of the necessity to compute the Hessian matrix and solve the corresponding system of equations.

$$\boldsymbol{h}_{k+1} = \boldsymbol{h}_k - [\nabla^2 \mathcal{J}(\boldsymbol{h}_k)]^{-1} \nabla \mathcal{J}(\boldsymbol{h}_k) \ ,$$

where $\nabla^2 \mathcal{J}(\boldsymbol{h})$ is the Hessian matrix of $\mathcal{J}(\boldsymbol{h})$ with respect to $\boldsymbol{h}$. To compute the Hessian matrix we need the derivative of the $\mathrm{sign}(\cdot)$ function. Because $\mathrm{sign}(\cdot)$ have a discontinuity in the derivative at zero, we use a follows approximation :

$$\mathrm{sign}(x) \approx \tanh(\theta\, x) \ .$$

For large θ this approximates the true Generalized Gaussian prior while staying smooth around zero. Now, taking derivative of (19) with respect to $\boldsymbol{h}$ and tacking into the regularization parameter, we obtain the following expression of the Hessian matrix :

$$\nabla^2 \mathcal{J}(\boldsymbol{h}_k) = 2\ \boldsymbol{\mathcal{Q}}_M + \lambda\ \mathrm{diag}(\overline{\boldsymbol{h}}_k^{\varepsilon}) \ , \tag{24}$$

where

$$\overline{h}_k^{\varepsilon}(i) = p\ (\theta\ \cosh^{-2}(\theta\, h_k(i))\ |h_k(i)+\varepsilon|^{p-1} + (p-1)\ |h_k(i)+\varepsilon|^{p-2}) \qquad \text{for} \quad i = 1, \ldots, M(L+1) \ .$$

With the unit norm constraint, the update equation is given by :

$$\boldsymbol{h}_{k+1} = \frac{\boldsymbol{h}_k - \left(2\ \boldsymbol{\mathcal{Q}}_M + \lambda\ \mathrm{diag}(\overline{\boldsymbol{h}}_k^{\varepsilon})\right)^{-1}\left(2\ \boldsymbol{\mathcal{Q}}_M\boldsymbol{h}_k + \lambda\ \widetilde{\boldsymbol{h}}_k^{\varepsilon}\right)}{\left\|\boldsymbol{h}_k - \left(2\ \boldsymbol{\mathcal{Q}}_M + \lambda\ \mathrm{diag}(\overline{\boldsymbol{h}}_k^{\varepsilon})\right)^{-1}\left(2\ \boldsymbol{\mathcal{Q}}_M\boldsymbol{h}_k + \lambda\ \widetilde{\boldsymbol{h}}_k^{\varepsilon}\right)\right\|_2} \ .$$

D. Weighting parameter λ optimization

In order to optimize the weighting parameter λ, we propose to exploit the theoretical interpretation of this parameter given by the MAP approach. Indeed, by observing equation (15), we note that the weighting parameter λ is equal to the variance ration of the noise and the channel vector $\boldsymbol{h}$:

$$\lambda = \frac{2\,\sigma^2}{\beta^p} \ .$$

Therefore, we can estimate the parameter λ by estimating the variance σ^2 and β. For the noise variance estimation, we exploit the spatio-temporal diversity of SIMO system by using Akaike's information criterion (AIC) [19]. Let define the spatio-temporal vector:

$$\boldsymbol{x}_d(n) = [\boldsymbol{x}^T(n)\dots\boldsymbol{x}^T(n-d+1)]^T = \boldsymbol{H}\,\boldsymbol{s}_d(n) + \boldsymbol{w}_d(t) \tag{25}$$

where $\boldsymbol{H}$ is block-Sylvester matrix of size $M\,d \times (d+L+1)$ and $\boldsymbol{s}_d(t) \stackrel{\text{def}}{=} [s(n)\dots s(n-L-1-d)]^T$, where d is a chosen processing window size. Under the data model assumption and for large window sizes (see [19] for more details), matrix $\boldsymbol{H}$ are full column rank. Hence, in the noiseless case, the rank of the data covariance matrix $\boldsymbol{R}_x \stackrel{\text{def}}{=} \mathbb{E}\left[\boldsymbol{x}_d(n)\boldsymbol{x}_d^H(n)\right]$ is equal to $d+L+1$ which corresponds to the dimension of the signal subspace.

Therefor, our approach consists in estimating the rank of the sample3 averaged covariance matrix $\boldsymbol{R}_x$ in order to estimate the dimension of the signal subspace and estimating the noise variance by meaning the eigenvalues of the averaged covariance matrix corresponding to the noise subspace. The estimation of the rank value is done here by Akaike's criterion [19] according to:

$$r = \arg\min_k \left[-2\log\left(\frac{\prod_{i=k+1}^{Md} \gamma_i^{1/(Md-k)}}{\frac{1}{Md-k}\sum_{i=k+1}^{Md}\gamma_i} \right)^{(Md-k)T} + 2k(2Md-k) \right] \tag{26}$$

where $\gamma_1 \geq \ldots \geq \gamma_{Md}$ represent the eigenvalues of $\boldsymbol{R}_x$. Note that it is not necessary at this stage, to know exactly the channel degree L as long as $d > (L+1)$ (i.e. an over-estimation of the channel degree is sufficient). Finally, the noise variance is estimating as follows:

$$\widehat{\sigma}^2 = \frac{1}{Md - r} \sum_{k=r+1}^{Md} \gamma_k \ .$$

Now, for the parameter λ optimization, we propose an iterative procedure, where the λ is compute at each iteration by:

$$\lambda_k = \frac{2\,\widehat{\sigma}^2}{\widehat{\beta}_k^p} \tag{27}$$

where $\widehat{\beta}_k$ is estimated at each iteration as the empirical standard deviation of the vector $\boldsymbol{h}_k$:

$$\widehat{\beta}_k = \sqrt{\frac{\Gamma(\frac{1}{p})}{\Gamma(\frac{3}{p})\, M\,(L+1)} \sum_{i=1}^{M(L+1)} |h_k(i)|^2}$$

E. Discussion

In this section, we discuss about the choice of the parameter p and his impact on the optimization strategy. We observe from (17) that the parameter p express the sparsity degree of the desired vector $\boldsymbol{h}$. In other words, when $p \to 0$ the criterion optimization will be equivalent to solving a combinatorial problem associated to ℓ_0 norm which is a counting measure of the nonzero elements of the vector $\boldsymbol{h}$. Unfortunately, it must be noted that (17) is a non-convex optimization problem when $0 < p < 1$ and all of the algorithms considered her are only designed to produce local minimum. However, in what follows we propose a kind of method based on the algorithms presented above to find a solution of the non-convex criterion.

a) $p = 1$: In this case criterion (17) is convex. Therefor, the proposed algorithms may be used without the regularization parameter ($\varepsilon = 0$).

b) $p = 0$: In this case the criterion will be convex if we introduce the regularization parameter ($\varepsilon > 0$). This assertion is confirmed by Candes et al. work's [20], where the case $p = 0$ with regularization

parameter is equivalent to reweighted ℓ_1 criterion. In this particular case, we can exploit the ε optimization solution proposed in [20]. The parameter ε is adapted at each iteration as a function of the vector $\boldsymbol{h}_k$. Let $|h_k|_{(i)}$ denote a reordering of $|h_k(i)|$ in decreasing order of magnitude. Then,

$$\varepsilon = \max\left\{ |h_k|_{(i_0)}, 10^{-3} \right\}$$

where $i_0 = \left\lceil \frac{1}{4} \frac{M(L+1)}{\log(T-L-1)-\log(M(L+1))} \right\rceil$

c) $0 < p < 1$*:* In this case the criterion (17) is non-convex. One kind of solution based on the presented algorithms, consists on a two step optimization approach, described as follows:

Step 1: Perform an optimization with $p = 1$ or $p = 0$ to ensure the convexity of the criterion and apply the presented optimization algorithms. An alternative solution is to find the eigenvector associated to the smallest eigenvalue of the matrix $\boldsymbol{\mathcal{Q}}$ (the solution without the ℓ_p penalty term).

Step 2: Select p in the range $]0, 1[$ and initialize the optimization algorithm by the solution found in step 1. Thus, we expect that the initialization will be nearest to the global minimum and then avoid the local minima problem.

An alternative procedure is to use Mosek-based interior-point linear programming solver or any other efficient linear programming solver.

V. Adaptive implementation

For blind channel identification to be practically useful in real-time applications, it is imperative that the algorithm should be computationally simple and can be adaptively implemented. In this section, we present an adaptive implementation of the CR and penalized CR algorithms with least mean square (LMS), normalized least mean square (NLMS) and proportional normalized least square (PNLMS) approaches.

A. Adaptive CR implementation

In the same way that in the block approach, the cross-relations between the sensor outputs can be exploited to estimate the channel impulse responses. In this case, we can rewrite the CR criterion as

follows:

$$\overline{\mathcal{J}}(\boldsymbol{h}) = \boldsymbol{h}^H \overline{\boldsymbol{\mathcal{Q}}}(n) \boldsymbol{h}\,, \tag{28}$$

where

$$\overline{\boldsymbol{\mathcal{Q}}}(n) = \gamma\, \overline{\boldsymbol{\mathcal{Q}}}(n-1) + \overline{\boldsymbol{\mathcal{X}}}_M^H(n) \overline{\boldsymbol{\mathcal{X}}}_M(n)\ ,$$

with $\overline{\boldsymbol{\mathcal{X}}}_M(n)$ given computed by using equations (7), (8) and

$$\overline{\boldsymbol{\mathcal{X}}}_2(n) = [\overline{\boldsymbol{x}}_2(n),\ -\overline{\boldsymbol{x}}_1(n)] \tag{29}$$

and

$$\overline{\boldsymbol{\mathcal{X}}}_l(n) = \left[\begin{array}{ccc|c} & \overline{\boldsymbol{\mathcal{X}}}_{l-1}(n) & & \mathbf{0} \\ \hline \overline{\boldsymbol{x}}_l(n) & & \mathbf{0} & -\overline{\boldsymbol{x}}_1(n) \\ & \ddots & & \vdots \\ \mathbf{0} & & \overline{\boldsymbol{x}}_l(n) & -\overline{\boldsymbol{x}}_{l-1}(n) \end{array}\right] \tag{30}$$

with $l = 3, \ldots, M$ and :

$$\overline{\boldsymbol{x}}_l(n) = \left[\begin{array}{ccc} x_l(n) & \ldots & x_l(n-L) \end{array}\right]\ . \tag{31}$$

Therefore, the desired solution for $\boldsymbol{h}$ is determined by minimizing the mean value of the cost function $\overline{\mathcal{J}}(\boldsymbol{h})$:

$$\widehat{\boldsymbol{h}} = \underset{\|\boldsymbol{h}\|_2=1}{\arg\min} \left\{ \mathbb{E}\left[\boldsymbol{h}^H \overline{\boldsymbol{\mathcal{Q}}}(n) \boldsymbol{h} \right] \right\}$$

Here, an LMS algorithm is proposed to solve this minimization problem efficiently. Then, the filter coefficient vector is then update by:

$$\boldsymbol{h}_{n+1} = \frac{\boldsymbol{h}_n - 2\,\mu\, \overline{\boldsymbol{\mathcal{Q}}}(n)\, \boldsymbol{h}_n}{\|\boldsymbol{h}_n - 2\,\mu\, \overline{\boldsymbol{\mathcal{Q}}}(n)\, \boldsymbol{h}_n\|_2}\ .$$

However, the LMS algorithm suffers from slow and data-dependent convergence behavior. The normalized LMS (NLMS) [21], an equally simple, but more robust variant of the LMS algorithm, exhibits a better

balance between simplicity and performance than the LMS algorithm. Therefore, we present in what follow, a NLMS approach to optimize the CR criterion. One easy way to find adaptive algorithms that adjust the new channel vector $\boldsymbol{h}_{n+1}$ from the old one $\boldsymbol{h}_n$ is to minimize the following function [21]:

$$\mathcal{L}\left[\boldsymbol{h}_{n+1}\right] = d\left[\boldsymbol{h}_{n+1}, \boldsymbol{h}_n\right] \, + \, \mu \, \boldsymbol{h}_{n+1}^H \overline{\boldsymbol{\mathcal{Q}}}(n) \boldsymbol{h}_{n+1}$$

where $d\left[\boldsymbol{h}_{n+1}, \boldsymbol{h}_n\right]$ is a measure of distance from the old to the new channel vector and η is a positive constant. The magnitude of represents the importance of correctiveness compared to the importance of conservativeness [3]. To minimize $\mathcal{L}\left[\boldsymbol{h}_{n+1}\right]$, we need to set its derivative $\frac{\partial \mathcal{L}[\boldsymbol{h}_{n+1}]}{\partial \boldsymbol{h}_{n+1}}$ to zero. Hence, the solution will be found by solving the equation:

$$\frac{d\left[\boldsymbol{h}_{n+1}, \boldsymbol{h}_n\right]}{\partial \boldsymbol{h}_{n+1}} \, + \, 2\mu \, \overline{\boldsymbol{\mathcal{Q}}}(n) \boldsymbol{h}_{n+1} = \boldsymbol{0} \tag{32}$$

The LMS algorithm is easily obtained from (32) by using the squared Euclidean distance

$$d\left[\boldsymbol{h}_{n+1}, \boldsymbol{h}_n\right] = \|\boldsymbol{h}_{n+1} - \boldsymbol{h}_n\|_2^2 \tag{33}$$

so that and from equation (32), we obtain that:

$$\boldsymbol{h}_{n+1} = \left(\boldsymbol{I} + \mu \overline{\boldsymbol{\mathcal{Q}}}(n)\right)^{-1} \boldsymbol{h}_n \ , \tag{34}$$

but, according to the constraint:

$$\boldsymbol{h}_n^H \left(\boldsymbol{I} + \mu \overline{\boldsymbol{\mathcal{Q}}}(n)\right)^{-1} \overline{\boldsymbol{\mathcal{Q}}}(n) \left(\boldsymbol{I} + \mu \overline{\boldsymbol{\mathcal{Q}}}(n)\right)^{-1} \boldsymbol{h}_n = 0 \ . \tag{35}$$

By using the first order approximation of the inverse matrix, the equation (42) will be:

$$\boldsymbol{h}_n^H \left(\boldsymbol{I} - \mu \overline{\boldsymbol{\mathcal{Q}}}(n)\right) \overline{\boldsymbol{\mathcal{Q}}}(n) \left(\boldsymbol{I} - \mu \overline{\boldsymbol{\mathcal{Q}}}(n)\right) \boldsymbol{h}_n = 0 \ . \tag{36}$$

Therefor, the optimal step size μ is determined as a positive solution of the second order equation:

$$\mu^2 - 2a\mu + b = 0$$

where

$$a = \frac{\boldsymbol{h}_n^H \overline{\boldsymbol{\mathcal{Q}}}(n)^2 \boldsymbol{h}_n}{\boldsymbol{h}_n^H \overline{\boldsymbol{\mathcal{Q}}}(n)^3 \boldsymbol{h}_n}$$

and

$$b = \frac{\boldsymbol{h}_n^H \overline{\boldsymbol{\mathcal{Q}}}(n) \boldsymbol{h}_n}{\boldsymbol{h}_n^H \overline{\boldsymbol{\mathcal{Q}}}(n)^3 \boldsymbol{h}_n}$$

B. Adaptive sparse CR implementation

In the same way that in the adaptive standard CR implementation, the rewriting of the sparse CR criterion in adaptive case, leads to the following criterion:

$$\overline{\mathcal{J}}(\boldsymbol{h}) = \boldsymbol{h}^H \overline{\boldsymbol{\mathcal{Q}}}(n) \boldsymbol{h} \, + \lambda \|\boldsymbol{h}\|_p^p \, , \tag{37}$$

Therefore, the desired solution for $\boldsymbol{h}$ is determined by minimizing the mean value of this cost function and the LMS solution is given by the following adaptive solution:

$$\boldsymbol{h}_{n+1} = \frac{\boldsymbol{h}_n - \mu \left(2 \, \overline{\boldsymbol{\mathcal{Q}}}(n) \, \boldsymbol{h}_n + \lambda \, \widetilde{\boldsymbol{h}}_n^{\varepsilon} \right)}{\left\| \boldsymbol{h}_n - \mu \left(2 \, \overline{\boldsymbol{\mathcal{Q}}}(n) \, \boldsymbol{h}_n + \lambda \, \widetilde{\boldsymbol{h}}_n^{\varepsilon} \right) \right\|_2} \, . \tag{38}$$

where

$$\widetilde{h}_n^{\varepsilon}(i) = \operatorname{sign}\left(h_n(i)\right) \; \left(|h_n(i)| + \varepsilon\right)^{p-1} \qquad \text{for} \quad i = 1, \ldots, M(L+1) \; , \tag{39}$$

In this paper we consider only the case where $(p, \varepsilon) \in \left\{ (1,0), (0, \varepsilon_0) \right\}$ to avoid the non-convexity problem of the proposed criterion.

For the NLMS approach in the case of sparse CR criterion, we assume the same development as in section V-A . By tackint into a count the sparsity penalty term, the new cost function can be express as:

$$\mathcal{L}\left[\boldsymbol{h}_{n+1}\right] = d\left[\boldsymbol{h}_{n+1}, \boldsymbol{h}_n\right] \, + \, \mu \, \boldsymbol{h}_{n+1}^H \overline{\boldsymbol{\mathcal{Q}}}(n) \boldsymbol{h}_{n+1} \, + \, \lambda \|\boldsymbol{h}_{n+1}\|_p^p$$

In the same way that shown is section V-A , in order to minimize $\mathcal{L}\left[\boldsymbol{h}_{n+1}\right]$ in the case squared Euclidean distance, we need to set its derivative $\frac{\partial \mathcal{L}[\boldsymbol{h}_{n+1}]}{\partial \boldsymbol{h}_{n+1}}$ to zero. Hence, the solution will be found by solving the equation

$$2 \, \left(\boldsymbol{h}_{n+1} - \boldsymbol{h}_n \right) \, + \, 2\mu \, \overline{\boldsymbol{\mathcal{Q}}}(n) \boldsymbol{h}_{n+1}) \, + \, \lambda \widetilde{\boldsymbol{h}}_{n+1} = \boldsymbol{0} \tag{40}$$

To avoid the divergence problem of the derivative of the ℓ_p norm, we replace the $\widetilde{\boldsymbol{h}}_{n+1}$ by $\widetilde{\boldsymbol{h}}^{\varepsilon}_{n+1}$ such as is define by equation (), and by using the following approximation $\widetilde{\boldsymbol{h}}^{\varepsilon}_{n+1} \approx \widetilde{\boldsymbol{h}}^{\varepsilon}_{n}$ for $(p,\varepsilon) \in \left\{(1,0),(0,\varepsilon_0)\right\}$, we obtain that:

$$\boldsymbol{h}_{n+1} = \left(\boldsymbol{I} + \mu\overline{\boldsymbol{\mathcal{Q}}}(n)\right)^{-1}\left(\boldsymbol{h}_n - \frac{\lambda}{2}\widetilde{\boldsymbol{h}}^{\varepsilon}_{n}\right) \ , \tag{41}$$

and according to the CR constraint ;

$$\left(\boldsymbol{h}_n - \frac{\lambda}{2}\widetilde{\boldsymbol{h}}^{\varepsilon}_{n}\right)^H \left(\boldsymbol{I} + \mu\overline{\boldsymbol{\mathcal{Q}}}(n)\right)^{-1}\overline{\boldsymbol{\mathcal{Q}}}(n)\left(\boldsymbol{I} + \mu\overline{\boldsymbol{\mathcal{Q}}}(n)\right)^{-1}\left(\boldsymbol{h}_n - \frac{\lambda}{2}\widetilde{\boldsymbol{h}}^{\varepsilon}_{n}\right) = 0 \ . \tag{42}$$

and we can find easely from the previous section that the optimal step size μ is a solution of

$$\mu^2 - 2\widetilde{a}\mu + \widetilde{b} = 0$$

where

$$\widetilde{a} = \frac{\left(\boldsymbol{h}_n - \frac{\lambda}{2}\widetilde{\boldsymbol{h}}^{\varepsilon}_{n}\right)^H \overline{\boldsymbol{\mathcal{Q}}}(n)^2 \left(\boldsymbol{h}_n - \frac{\lambda}{2}\widetilde{\boldsymbol{h}}^{\varepsilon}_{n}\right)}{\left(\boldsymbol{h}_n - \frac{\lambda}{2}\widetilde{\boldsymbol{h}}^{\varepsilon}_{n}\right)^H \overline{\boldsymbol{\mathcal{Q}}}(n)^3 \left(\boldsymbol{h}_n - \frac{\lambda}{2}\widetilde{\boldsymbol{h}}^{\varepsilon}_{n}\right)}$$

and

$$\widetilde{b} = \frac{\left(\boldsymbol{h}_n - \frac{\lambda}{2}\widetilde{\boldsymbol{h}}^{\varepsilon}_{n}\right)^H \overline{\boldsymbol{\mathcal{Q}}}(n) \left(\boldsymbol{h}_n - \frac{\lambda}{2}\widetilde{\boldsymbol{h}}^{\varepsilon}_{n}\right)}{\left(\boldsymbol{h}_n - \frac{\lambda}{2}\widetilde{\boldsymbol{h}}^{\varepsilon}_{n}\right)^H \overline{\boldsymbol{\mathcal{Q}}}(n)^3 \left(\boldsymbol{h}_n - \frac{\lambda}{2}\widetilde{\boldsymbol{h}}^{\varepsilon}_{n}\right)}$$

VI. SIMULATION

We present here some numerical simulations to assess the performance of the proposed algorithm. We consider a SIMO system with $M = 3$ outputs represented by polynomial transfer function of degree $L = 256$. The channel impulse response is a sparse sequence of random variables with Bernoulli-Gaussian distribution [13] :

$$f(h_i) = p_i\delta(h_i) + (1-p_i)\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(-h_i^2/2\sigma_i^2\right)$$

generated by the MATLAB function SPRANDN. We used the parameters $p_i = 0.5$ and $\sigma_i = 1$. The input signal is a 4-QAM i.i.d. sequence of length $T = 1024$. The observation is corrupted by addition white Gaussian noise with a variance σ^2 chosen such that the $SNR = \frac{\|\boldsymbol{h}\|^2}{\sigma^2}$ varies in the range $[5, 50]\ dB$. The

weighting parameter for the algorithm ISBSI is chosen as $\lambda = 1$. Statistics are evaluated over $N_r = 200$ Monte-Carlo runs and estimation performance are given by the normalized mean-square error criterion :

$$\begin{aligned} NMSE &= \frac{1}{N_r}\sum_{r=1}^{N_r}\min_{\alpha}\left(\frac{\|\alpha\widehat{\boldsymbol{h}}_r - \boldsymbol{h}\|^2}{\|\boldsymbol{h}\|^2}\right) \\ &= \frac{1}{N_r}\sum_{r=1}^{N_r} 1 - \left(\frac{\widehat{\boldsymbol{h}}_r^H \boldsymbol{h}}{\|\widehat{\boldsymbol{h}}_r\|\|\boldsymbol{h}\|}\right)^2 , \end{aligned}$$

where $\widehat{\boldsymbol{h}}_r$ denotes the estimated channel coefficient vector at the r^{th} Monte-Carlo run and α is a scalar factor that compensates for the scale indeterminacy of the BSI problem.

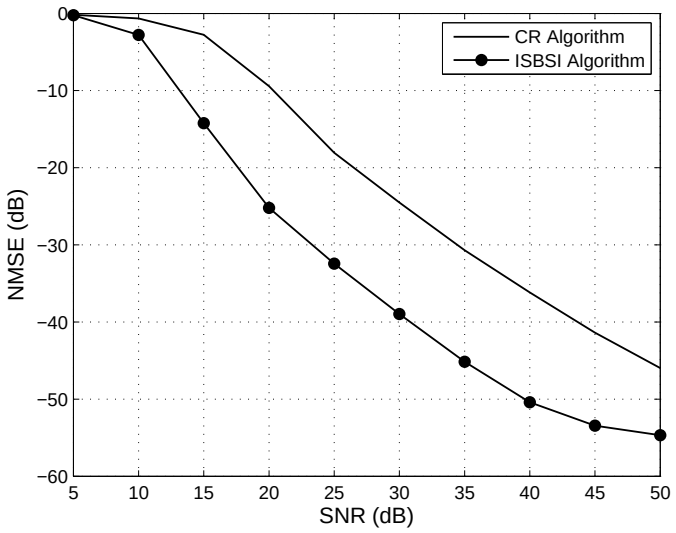

Fig. 1. Normalized mean-square error (NMSE) versus the SNR for SIMO system with 3 sensors: comparison between CR and the proposed ISBSI algorithm.

In figures 1 and 2, the normalized mean-square error is plotted versus the SNR for the proposed ISBSI algorithm and the CR and CS algorithms respectively. It is clearly shown that our algorithm (ISBSI) performs better in terms of the normalized mean-square error especially for moderate and high SNR.
In figure 3, the normalized mean-square error is plotted versus the SNR for the proposed MAP algorithm and the ML algorithm. We observe a same conclusion as in the previous figure.

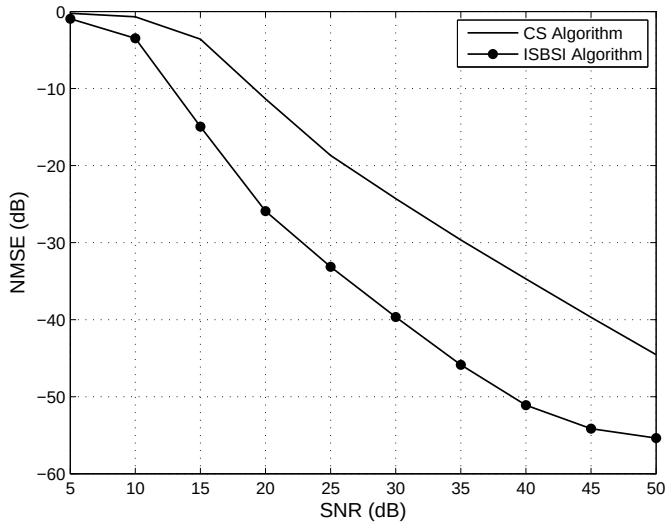

Fig. 2. Normalized mean-square error (NMSE) versus the SNR for SIMO system with 3 sensors: comparison between CS and the proposed ISBSI algorithm.

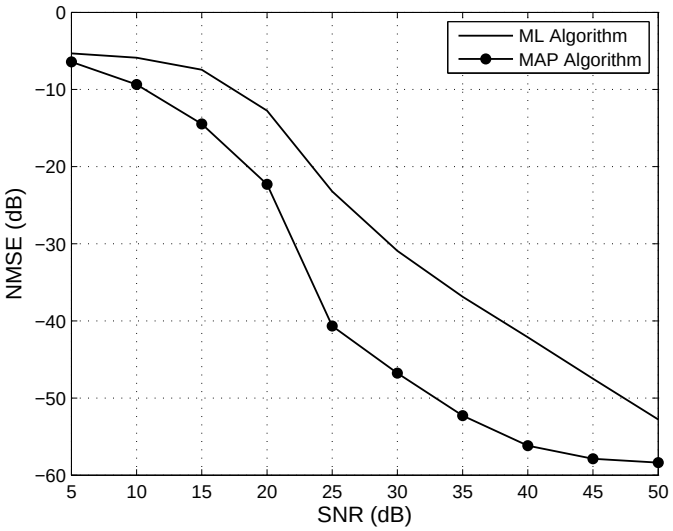

Fig. 3. Normalized mean-square error (NMSE) versus the SNR for SIMO system with 3 sensors: comparison between ML and the proposed MAP algorithm.

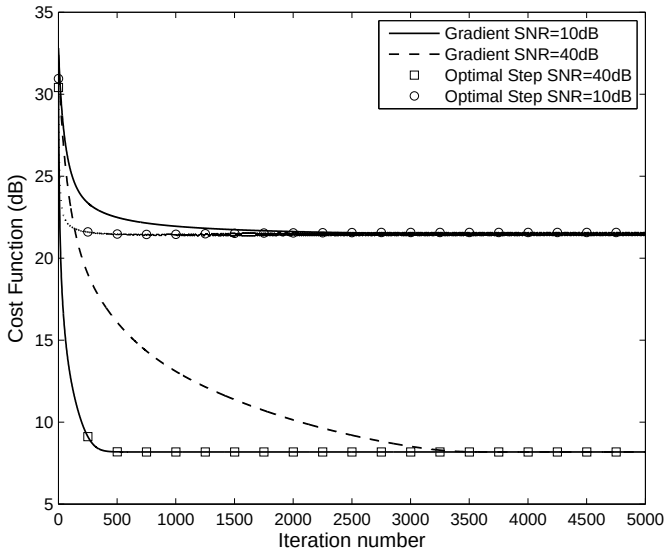

Fig. 4. Evolution of the cost function in dB as a function of the iteration number for SIMO system with 3 sensors.

In figures 4 and 5, we represent the evolution of the cost function in dB as a function of the iteration number for the gradient with fixed and optimal step size and Newton techniques. It is shown that the Newton technique converges much faster than the optimal and fixed step size one.

In figure 6, we represent the evolution of the NMSE in dB as a function of the overestimated channel order for the ISBSI algorithm. This figure illustrates the robustness of our algorithm against channel order overestimation errors.

In figure 7, we represent the NMSE as a function of the SNR for different values of the weighting parameter λ. We observe that, for large SNRs, small λ values are preferred, while for low SNRs, the large λ values are those leading to the best channel estimation accuracy. From this observation we plan for our futur works to study the optimization of the parameter λ in the ISBSI algorithm.

The plot in figure 8 presents the identification performance (NMSE) as function of the SNR for CR algorithm, ISBSI algorithm with $\lambda = 1$ and ISBSI algorithm with weighting parameter λ optimization technique. We observe that we increase the identification performance especially at low and moderate

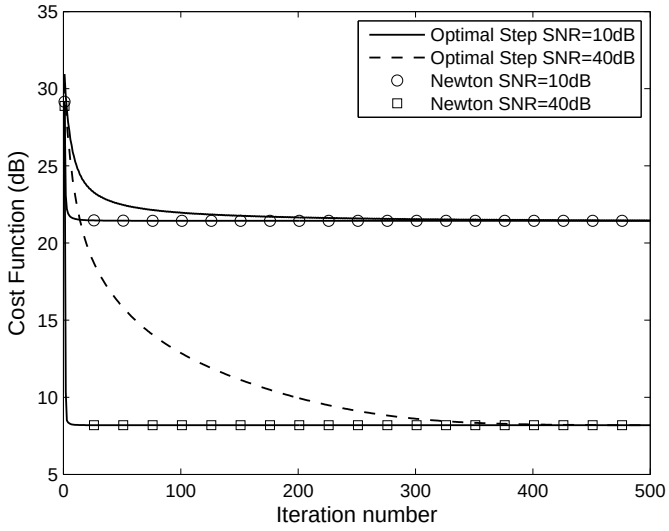

Fig. 5. Evolution of the cost function in dB as a function of the iteration number for SIMO system with 3 sensors.

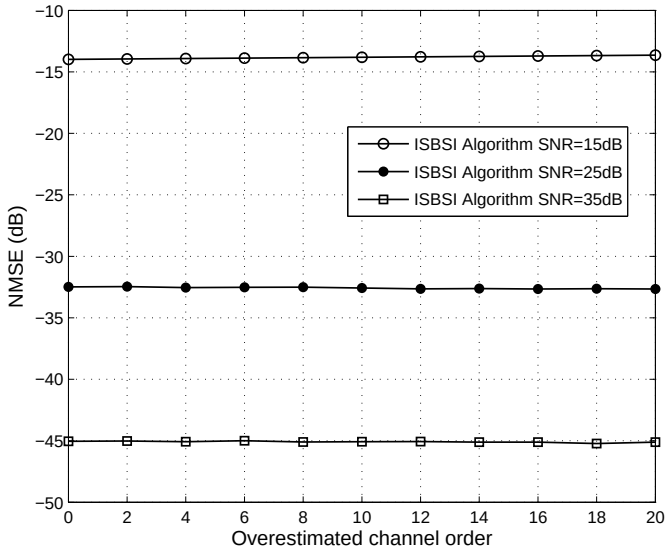

Fig. 6. Normalized mean-square error (NMSE) versus the overestimated channel order for SIMO system with 3 sensors and for different value of the SNR.

SNRs values by using the weighting parameter optimization.

In figure 9, we represent the identification performance of CR and ISBSI algorithms for reel channel

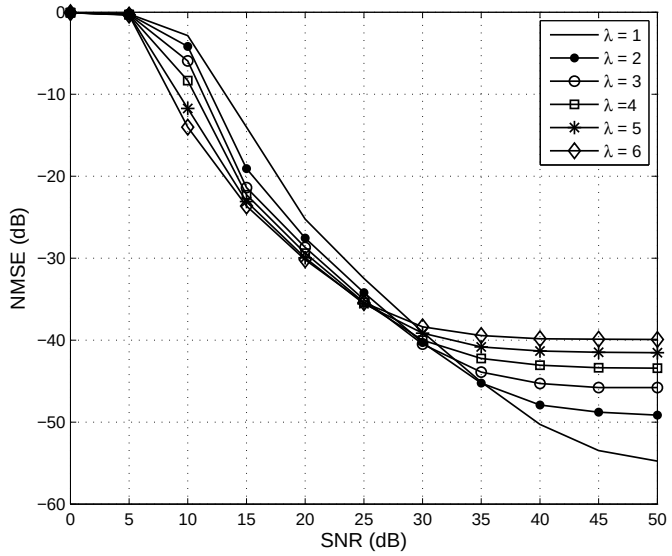

Fig. 7. Normalized mean-square error (NMSE) versus the SNR for SIMO system with 3 sensors: performance of our ISBSI algorithm for different value of the regularization parameter λ.

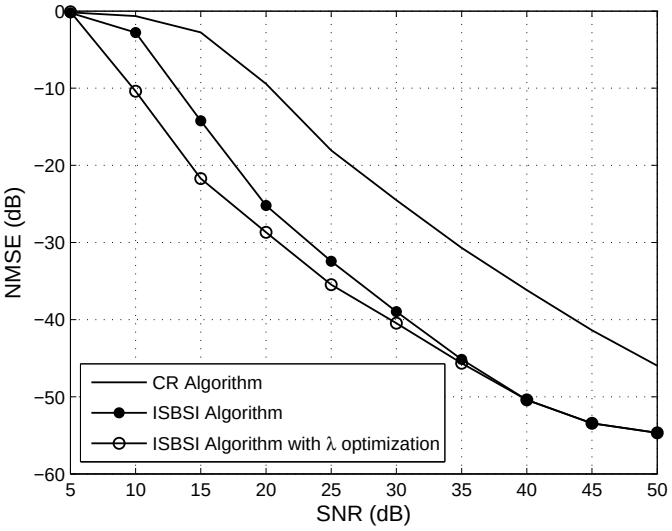

Fig. 8. Normalized mean-square error (NMSE) versus the SNR for SIMO system with 3 sensors: comparison between CR algorithm, ISBSI algorithm and ISBSI algorithm with λ optimization.

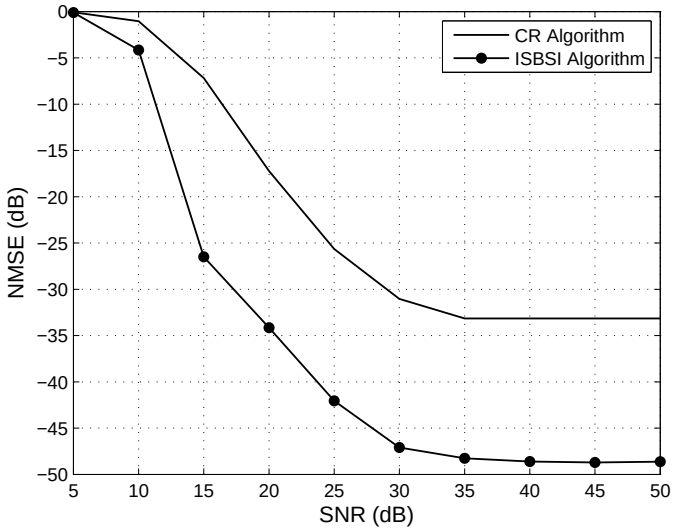

Fig. 9. Normalized mean-square error (NMSE) versus the SNR for SIMO system with 4 sensors with reel channel impulse response : comparison between CR and the proposed ISBSI algorithm.

impulse response. This impulse response has been measured between an Autonomous Underwater Vehicle (AUV) and 4 sensors placed on a ship. The ship is adrift. The AUV transmit to the ship and the AUV is diving of 15 m. The distance between the AUV And the ship is about 400 m. The acoustic signal used to the measure was at frequency 6.9 KHz with carrier frequency 35.4 KHz.

VII. Conclusion

This paper introduces an ISBSI algorithm and generalized version of the ML method for the blind estimation of sparse and long SIMO channel impulse responses. In the proposed methods, we use a channel sparsity measure together with the classical BSI criterion to improve the estimation quality and to take into account the sparsity of the channel. A gradient type technique with optimized step size and Newton technique has been considered for the optimization of the proposed cost function. Besides its improved performance, the new BSI method is robust against channel order overestimation errors. [22]–[25]

REFERENCES

[1] J. A. Cadzow, "Blind deconvolution via cumulant extrema," *IEEE Signal Processing Magazine*, vol. 6, no. 13, pp. 24–42, May 1996.

[2] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1310–1322, August 1997.

[3] L. Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1951–1968, October 1998.

[4] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982–2993, December 1995.

[5] W. Schreiber, "Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions," *Proceedings of the IEEE*, vol. 83, no. 6, pp. 958–981, June 1995.

[6] S. Ariyavisitakul, N. Sollenberger, and L. Greenstein, "Tap selectable decision-feedback equalization," *IEEE Transactions on Communications*, vol. 45, no. 12, pp. 1497–1500, December 1997.

[7] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 374–377, March 2002.

[8] J. Homer, I. Mareels, R. R. Bitmead, B. Wahlberg, and F. Gustafsson, "LMS estimation via structural detection," *IEEE Transactions on Signal Processing*, vol. 46, no. 10, pp. 2651–2663, October 1998.

[9] J. Homer, "Detection guided lms estimation of sparse channels," in *Proc. IEEE Globecom*, vol. 6, November 1998, pp. 3704–3709.

[10] M. Kocic, D. Brady, and M. Stojanovic, "Sparse equalization for real-time digital underwater acoustic communications," in *Proc. OCEANS*, vol. 3, San Diego, USA, October 1995, pp. 1417–1422.

[11] M. Stojanovic, "Recent advances in high-speed underwater acoustic communications," *IEEE Journal of Oceanic Engineering*, vol. 21, no. 2, pp. 125–136, April 1996.

[12] ——, "Retrofocusing techniques for high rate acoustic communications," *Journal of Acoustical Society of America*, vol. 117, no. 3, pp. 1173–1185, March 2005.

[13] M. Zibulevsky and Y. Zeevi, "Extraction of a source from multichannel data using sparse decomposition," *Neurocomputing*, vol. 49, no. 1, pp. 163–173, December 2002.

[14] A. Ossadtchi and S. Kadambe, "Over-complete blind source separation by applying sparse decomposition and information theoretic based probabilistic approach," in *Proc. ICASSP*, vol. 5, 2001, pp. 2801–2804.

[15] A. Hua and M. Wax, "Strict identifiability of multiple FIR channels driven by an unknown arbitrary sequence," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 756–759, March 1996.

[16] K. Abed-Meraim and E. Moulines, "A maximum likelihood solution to blind identification of multichannel FIR filters," in *Proc. EUSIPCO*, vol. 2, Edinburgh, Scotland, 1994, pp. 1011–1014.

[17] D. Starer and A. Nehorai, "Passive localization of near-field sources by path following," *IEEE Transactions on Signal Processing*, vol. 42, no. 3, pp. 677–680, March 1994.

[18] A. Aïssa-El-Bey, M. Grebici, K. Abed-Meraim, and A. Belouchrani, "Blind system identification using cross-relation methods: further results and developments," in *Proc.* 7^{th} *International Symposium on Signal Processing and Its Applications ISSPA*, vol. 1, Paris, France, July 2003, pp. 649–652.

[19] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. ASSP-33, no. 2, pp. 387–392, April 1985.

[20] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, December 2008.

[21] S. Haykin, Ed., *Adaptive filter theory*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[22] S. M. Aziz-Sbaï, A. Aïssa-El-Bey, and D. Pastor, "Contribution of statistical tests to sparseness-based blind source separation," *EURASIP journal on applied signal processing*, vol. 2012, no. 169, August 2012.

[23] A. Aïssa-El-Bey, K. Abed-Meraim, and C. Laot, "Adaptive blind estimation of sparse SIMO channels," in *7th International Workshop on Systems, Signal Processing and their Applications WOSSPA*, Tipaza, Algeria, May 2011.

[24] A. Kammoun, A. Aïssa-El-Bey, K. Abed-Meraim, and S. Affes, "Robustness of blind subspace based techniques using lp quasi-norms," in *11th IEEE International Workshop on Signal Processing Advances in Wireless Communications SPAWC*, Marrakech, Morocco, June 2010.

[25] F.-X. Socheleau, D. Pastor, and A. Aïssa-El-Bey, "Robust statistics based noise variance estimation : application to wideband interception of non-cooperative communications," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 1, pp. 746 – 755, January 2011.